ON COLOR SPACES FOR CYTOLOGY

Amalka Pinidiyaarachchi¹,² and Carolina Wählby¹

1. Centre for Image Analysis, Uppsala University, Sweden
2. University of Peradeniya, Sri Lanka

ABSTRACT

Quantitative measurements of stained nuclei of both cytological and histological samples are used in aid of decision making in many biomedical applications. The digital images produced in such processes are often in color, and segmentation methods of such images have been widely studied. This study compares the use of a number of different color spaces in the segmentation process without using priori information regarding the final segmentation result. The color space that gives the best results is then used in the process to obtain an initial segmentation result. The result is further processed using seeded watershed segmentation. The biological goal in this study is to use the methods effectively in extracting various features of cell nuclei that can be used in an expert decision making process.

1. INTRODUCTION

Cytology methods are important in the diagnosis of many diseases. Among quantitative parameters such as nuclear size, cell size, nuclear shape, nuclear-to-cytoplasmic ratio, cell shape and the nuclear discontinuity, the nuclear size is a major factor which enables the detection of malignant cells. Proper recognition of normal and abnormal cells seen in smears and the ability to assess them separately is an important factor in correct and prompt diagnosis.

Thyroid nodular lesions are one such common clinical problem where the quantitative measures mentioned above are used in decision making. Fine needle aspirations are used for the preliminary diagnosis before employing more invasive procedures such as surgical excision [1]. A correct and confident diagnosis at this stage is of great importance in planning further management strategies. If a surgical excision is carried out histological samples might be further tested for confirmation.

The initial diagnosis is usually done through microscopic examination of fine needle aspiration cytology smears stained to highlight the nucleus and cytoplasm. All conventional examinations are done through the microscope, and this requires the trained expert to examine all the cells in the smears visually. This is undoubtedly a tedious and time consuming process.

Computerized cell image analysis methods have been in practice over a number of decades [2]. Digital image processing tools that perform estimation of morphological parameters of individual cells have been used in various studies. However, it is difficult to use generalized analysis tools in the diagnosis of diseases. Even though some commercial image processing software have been developed, their high cost and inability to be customized for local requirements justify the development of a tool that uses digital image processing based techniques for display and analysis of microscopic images specific to a given area of pathology.

Assessment of nuclear size is usually subjective and not very accurate, especially when variations of nuclear size are subtle between different cells. The routine method of analyzing fine needle aspirates is to obtain a smear and stain using the Papanicolaou stain, and observe through the microscope. No measurements are obtained but decision is mainly based on the visual perception of the comparative size of each nucleus, combined with experience of the examiner. As opposed to this conventional approach, digital image processing techniques are expected to provide a much more robust and efficient means of analysis.

2. MATERIALS AND METHODS

2.1 Specimen preparation and image acquisition

Fine needle aspiration cytology smears of thyroid lesions stained with the Papanicolaou stain were used in the study and the images were acquired using an Olympus BX 50 microscope with an Olympus DP12 digital camera attached and 40x objective magnification and a 10x eyepiece magnification.

2.2 Color space selection

When compared to gray scale images, color provides information in addition to intensity and this information can sometimes be useful or even necessary in image analysis tasks. Most gray level image segmentation techniques can be extended to color images as well. The methods such as histogram thresholding, edge detection, fuzzy approaches etc. can be applied to each component of the color image separately and combined to obtain a final result. However, when the image is separated to the three color components R, G and B, the color information is scattered and the information that humans can perceive is lost.

Use of various color spaces in automatic quantification of immunohistochemistry has been done in many
instances [3]. Since color in images provide additional information directly attached to regions within the images, it is desirable to use color information in the segmentation process.

The selection of an appropriate color space thus becomes an important issue since the final result might exceedingly depend upon the choice. In many instances additional color spaces other than the well known ones (RGB, HSI, LUV etc) have been introduced with the aim of obtaining maximum amount of information that is attached to color. However, the selection might be heavily dependent on the type of application as well as the images available. Generalization of one color space usage to widely varying types of images is nearly impractical.

A numerical measure that checks the usability of different color spaces is proposed in [4]. This method extracts objects from sample images by manual criteria and use the information in determining which color space best corresponds to the transition between regions. A maximum value of the measure ‘color space choice’ (CSC) is said to be the requirement for the choice of a particular color space. Attempts have also been made to automate such selection procedures. Some of these automations test the segmentation results given the good segmentation satisfying a number of requirements such as regions being homogeneous, without holes, significantly different from adjacent regions, with smooth boundaries etc [5]. The selection of an appropriate color space depending on such criteria might yield good results in many situations. However, these methods suffer from a number of difficulties that might arise in the process such as the need of a ground truth and prior knowledge about the homogeneity of the color of interest etc. The methods might not give the expected result when larger numbers of regions are available. Extensive testing is necessary to make a proper decision on choice of parameters for automated methods [6].

Aiming to obtain a maximum amount of information from the color images at hand, we have done manual tests over a number of color spaces.

**RGB**

The first and the obvious choice was the RGB color space in spite of some of the well known drawbacks incorporated with it [7]. The original images are available as RGB and they were directly used for testing in this case.

**PCA**

The next choice was a principle component analysis (PCA) -based color space inspired by the ideas in [3]. PCA in is a well-known and widely used statistical technique for feature extraction and it has been used in image analysis as a method of image compression and sharpening as well [8]. The 3 principal components obtained by applying PCA to the RGB vector space of the images can be used separately in different image segmentations. For example the first principal component (PC1) can be used to segment dark objects from a light background since it represents the lightness of the image. The second and third components (PC2 and PC3) can be used for segmenting objects depending on color [3]. The

![Figure 1: Initial segmentation result on the four color spaces RGB, PCA, HSV and CSF.](image-url)
color spaces we have used for testing make use of all three components PC1, PC2 and PC3 to compose a PCA color space.

**HSV**

From the other common color spaces, we have also used the direct conversion from RGB to HSV color space. Criteria described in 2.3 have been used in obtaining an initial segmentation from the RGB, PCA and HSV images.

### 2.3 Initial segmentation in RGB, PCA and HSV

A user specified color region (a mask) which denotes the nucleus type of interest was used as a representative sample to initiate the segmentation, in order to obtain higher reliability of segmentation. The same sample was used for all three color spaces. The average color to be segmented was calculated using this sample. The distance to this mean color vector in the color space from each of the pixels in the image was calculated and as a selection criteria of similarity between the mean and each of the other pixels, a threshold value which is the standard deviation (SD) of the sample, i.e., the average distance from the mean to a data point in the sample was used. The distance measure was either as Euclidean or the generalized Mahalanobis distance that uses the covariance matrix of the selected sample representing the color of interest, in measuring the distance. If the segmentation based on this initial threshold is not satisfactory, a multiple of the initial value of the SD can be used until the result becomes satisfactory. A size thresholding which remove the small objects was applied to the results.

### 2.4 Color Space Fusion (CSF)

Instead of using classification in 3D color space, the color components can be fused into one image. One example of fusion of components of various color spaces is presented in [4]. It makes use of advantages attached to each of the fused components. The formula used here is:

\[
I_s = \frac{(I_{RGB}^C - I_{RGB}^G) + I_{Luv}}{2}
\]

where \(I_{RGB}^C\) denotes the color component \(C_1\) in C1C2C3 color space. The idea of this fusion is to use \(I_{Luv}\) to bring out the clear nuclei and to use the \((I_{RGB}^C - I_{RGB}^G)\) component to level the difference between the clear nuclei and the dark nuclei. The image \(I_s\) is expected to restrict intensity variation of nuclei to a small interval of values allowing an automatic threshold based on the intensity histogram to yield a good segmentation result.

### 2.5 Post processing

The initial segmentation result obtained here does not separate clusters of nuclei. Watersheds are a widely used method in many cell image segmentation procedures. It has been proven to be successful in a number of different situations with varying types of input information (seeds, then located at approximately the center of each nucleus. These seeds together with watershed segmentation successfully segments some of the clustered nuclei. A region merging step that removes all non-seeded regions reduces over segmentations in the result.

### 3. RESULTS AND DISCUSSION

The results of applying these methods to a set of images prepared as mentioned above can be seen in Figures 1 and 2. Figure 1 shows the set of original images (a), (b), (c), (d), (e), (f) and segmentation on RGB, PCA, HSV and CSF color space respectively in rows that follow. The results after seeded watershed are shown in Figure 2 for images (a), (b) and (c) in Figure 1 after initial segmentation using RGB color space.

When the CSC measure mentioned above was tested on nuclei segmentation of cytological color images approximately similar to images we have used it selects RGB as the choice of color space. This is justified by our results as long as a careful mask selection is done by the user.

All four methods yield satisfactory results in the initial segmentation step on image (a) in Figure 1. This can be considered as one of the simplest cases where we have clear dark nuclei in a light background. The CF result has the advantage over the rest in this case. It gives the nice initial result with automatic thresholding with no user interaction whereas the rest of the results need the initial mask selection to be given as an input. A careful selection of this mask is essential for the result to be satisfactory.
gradient images, distance transforms etc.\)[9,10]. In our implementation, seeds are automatically obtained from the initially segmented regions using the distance transform followed by the h-maxima transform. Seeds are

Thus if user interaction should be minimized and the samples have rather clear nuclei in a lighter background, the CSF method is preferred as an initial segmentation step.

Images with weak contrast and cell nuclei with internal structure of varying color are poorly segmented in all the tested color spaces. This will call for more user interaction to select only the well segmented and less clustered images for further processing to extract the required features.

The images containing cells with mainly isolated nuclei gives much nicer results compared to the other images. Segmentation of slightly clustered light nuclei in a lighter background is partly successful. Images where dark nuclei appear in dark background and in clusters are still an issue to be further addressed, see Figure 2. (c).

3.2 CONCLUSION AND FUTURE WORK

The methods discussed in this paper use color information available in the RGB, PCA and HSV color spaces to do an initial segmentation and create seeds that are given as input to the seeded watershed segmentation. When user input is available about the type of nuclei to be segmented (i.e. mask selection) and when the images have less clustered nuclei the segmentation gives satisfying results. One option at this stage is to have further user input to select a set of well segmented nuclei from the segmentation result and to obtain the required feature measures on the selected set only. This approach has the limitation of the call for more user input and also the risk of having only very few well segmented nuclei in the difficult cases limiting the reliability of the results obtained.

The color watershed used in [4] is said to obtain good results since color edge information is used in the watershed. It is desirable to test such methods on our images to see if improvements could be made to segment the heavily clustered nuclei.

Once satisfying results are obtained we will compare the results with the corresponding histological sample images obtained after surgical excision. The aim of this project is to test the reliability of the results obtained on the assessments made on the cytology samples. Running the test on a larger set of data from both cytology and histology would enable automation of the decision making process based on the measurements obtained.

Acknowledgments

The authors would like to thank Dr. N.V.I. Ratnatunga, Department of Pathology, University of Peradeniya Sri Lanka for providing the images and the first author is supported by SIDA (The Swedish International Development Agency).

References